

Klasterisasi Daerah Rawan Kriminalitas di Sulawesi Tenggara Menggunakan Metode *K-Means Clustering*

Muh. Afdal Ziqri Ramadhan¹, Rizal Adi Saputra^{2*}, Jumadil Nangi³

^{1,2,3} Jurusan Teknik Informatika Universitas Halu Oleo

Email: ¹muhafdalziqir2204@gmail.com, ^{2*}rizaladisaputra@uho.ac.id, ³jumadilnangi@uho.ac.id

Abstract— Crime is anything that relates to unlawful behavior. In every society, the presence of crime affects social dynamics, levels of security, and general well-being. Southeast Sulawesi in the midst of its diversity is facing a significant increase in crime, this is evident from the 3,000 cases that have been reported. The purpose of this research is to categorize crime-prone areas in Southeast Sulawesi using the K-Means Clustering method. K-Means is a research data analysis method or a data mining method that will carry out modeling without supervision (unsupervised). The data from this research was taken from the sultra.bps.go.id web page. The data taken in the form of data on the number of crimes and the percentage of crime victims in the regions of Southeast Sulawesi. The results of this study obtained 3 clusters of crime-prone areas in Southeast Sulawesi. The clustering evaluation results if calculated by the Davies Boulden Index method is as much as 0.5904 index.

Keywords— clustering, cluster, crime, k-means, southeast sulawesi.

Abstrak— Kriminalitas merupakan segala sesuatu yang berhubungan dengan perilaku yang melanggar hukum. Dalam setiap masyarakat, keberadaan kriminalitas mempengaruhi dinamika sosial, tingkat keamanan, dan kesejahteraan umum. Sulawesi Tenggara di tengah keragamannya menghadapi peningkatan signifikan akan kejahatan, hal ini terbukti dari 3.000 kasus yang telah dilaporkan. Tujuan dari Penelitian ini untuk mengkategorikan daerah rawan kriminalitas di Sulawesi tenggara menggunakan metode *K-Means Clustering*. *K-Means* ialah metode analisis data penelitian atau suatu metode data mining yang akan melaksanakan pemodelan tanpa adanya pengawasan (*unsupervised*). Data dari penelitian ini diambil dari halaman *website* sultra.bps.go.id. Data yang diambil berupa data jumlah kejahatan dan persentase korban kejahatan di daerah-daerah Sulawesi Tenggara. Hasil dari penelitian ini didapatkan 3 *cluster* daerah rawan kriminalitas di Sulawesi Tenggara. Adapun hasil evaluasi *clusteringnya* jika dihitung dengan metode *Davies Boulden Index* adalah sebanyak 0.5904 index.

Kata Kunci— cluster, clustering, kriminalitas, k-means, sulawesi tenggara.

I. PENDAHULUAN

Kriminalitas berasal dari istilah "kriminal" yang merujuk pada serangkaian tindakan atau perilaku yang terkait dengan pelanggaran hukum atau kejahatan yang dapat dikenai sanksi pidana menurut peraturan undang-undang. Istilah kriminalitas sendiri menggambarkan segala sesuatu yang berhubungan dengan perilaku yang melanggar hukum pidana atau perbuatan yang dianggap sebagai kejahatan. Dalam konteks ini, kriminalitas mencakup berbagai tindakan yang tergolong sebagai pelanggaran hukum atau perbuatan kriminal[1].

Dalam setiap masyarakat, keberadaan kriminalitas mempengaruhi dinamika sosial, tingkat keamanan, dan kesejahteraan umum. Sulawesi Tenggara sebagai sebuah wilayah yang beragam juga menghadapi tantangan serupa dalam mengelola masalah kriminalitas. Hal ini dibuktikan oleh data dari Badan Pusat Statistika (BPS) Sulawesi Tenggara yang mana total kriminalitas/kejahatan di Sulawesi Tenggara sejumlah 3.000 kasus [2].

Menghadapi tingkat kriminalitas yang semakin tinggi, penelitian diinisiasi dengan tujuan untuk melakukan pengelompokan daerah-daerah rawan kriminalitas di Sulawesi Tenggara. Metode yang dipilih dalam pengelompokan daerah ini adalah *K-Means Clustering*.

Metode ini dipilih karena kecenderungan untuk mengelompokkan data berdasarkan kesamaan karakteristik, sehingga dapat membantu identifikasi pola-pola kriminalitas yang mungkin tidak terlihat secara langsung [3].

Proses *K-Means Clustering* dimulai dengan menentukan jumlah kelompok (*cluster*) yang diinginkan. Kemudian, secara acak, titik pusat awal (*centroid*) untuk setiap kelompok ditetapkan. Data kemudian dikelompokkan berdasarkan kedekatan dengan *centroid* terdekat. Setelah itu, *centroid* dari setiap kelompok dihitung ulang berdasarkan rata-rata posisi anggota kelompok tersebut. Proses ini diulang hingga *centroid* tidak mengalami perubahan yang signifikan, dan kelompok-kelompok telah stabil [4].

Penerapan *K-Means Clustering* dalam konteks kriminalitas Sulawesi Tenggara dapat memberikan informasi mengenai pola-pola kriminalitas yang mungkin sulit dikenali secara manual. Dengan mengelompokkan daerah-daerah rawan kriminalitas, penelitian ini dapat membantu masyarakat agar lebih antisipasi terhadap tindak kejahatan.

Melalui penelitian ini diharapkan dapat memberikan kontribusi dalam pemahaman yang lebih baik terkait distribusi kriminalitas di Sulawesi Tenggara. Dengan hasil klasterisasi, diharapkan dapat diidentifikasi daerah-daerah

yang memerlukan perhatian khusus dalam upaya pencegahan dan penanggulangan kejahatan.

II. TINJAUAN PUSTAKA

A Metode Clustering

Clustering adalah salah satu metode dalam fungsionalitas data mining, di mana algoritma *clustering* digunakan untuk mengelompokkan data menjadi beberapa kelompok atau *cluster*. Dalam proses clustering, langkah untuk menentukan atau menggambarkan nilai kuantitatif dari tingkat kesamaan atau perbedaan antar data (*proximity measure*) menjadi kritis. Oleh karena itu, perbandingan beberapa metode yang umum digunakan seperti jarak *euclidean*, *manhattan*, dan *minkowski* menjadi suatu kebutuhan yang sangat penting. [4].

Hasil *clustering* yang efektif akan menciptakan tingkat keseragaman yang tinggi dalam suatu kelompok dan tingkat perbedaan yang rendah antar kelompok. Keseragaman diukur secara numerik terhadap dua objek, di mana nilai keseragaman antara dua objek akan meningkat seiring dengan tingginya kemiripan di antara keduanya. Sebaliknya, nilai keseragaman akan menurun jika objek yang dibandingkan memiliki perbedaan yang besar. Kualitas hasil clustering sangat tergantung pada metode yang digunakan. [5].

B Elbow Method

Elbow Method merupakan suatu teknik yang digunakan untuk menafsirkan dan menguji performa jumlah cluster yang optimal dengan memperhatikan nilai SSE (*Sum of Squared Errors*). Pada suatu titik tertentu dalam grafik, terjadi penurunan yang signifikan dengan adanya suatu lekukan yang disebut sebagai siku. Nilai tersebut kemudian dianggap sebagai jumlah *cluster* yang paling baik atau optimal (*k*). Untuk membandingkannya, SSE (*Sum of Square Error*) dihitung untuk setiap nilai *cluster*. Hal ini karena semakin besar jumlah cluster *K*, nilai SSE akan semakin kecil [6]. Adapun rumus untuk mencari SSE adalah sebagai berikut :

$$SSE = \sum_{k=1} \sum_{x_i \in S_k} X_i - C_k \quad (1)$$

Keterangan :

K = Jumlah klaster

X_i = Jumlah data

C_k = Jumlah clusteri pada cluster ke K

C K-Means

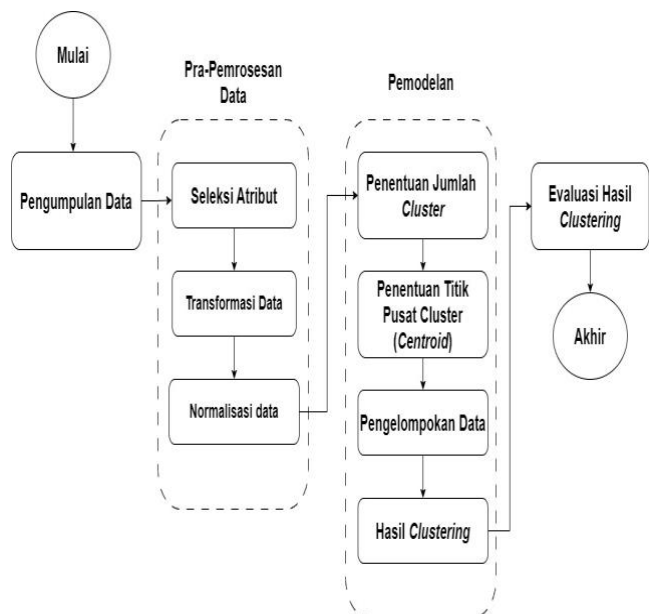
K-Means merupakan suatu teknik analisis data penelitian atau metode data mining yang dilakukan tanpa pengawasan (*unsupervised*) [7]. Pendekatan *K-Means* berusaha mengelompokkan data ke dalam kelompok-kelompok berdasarkan kemiripan fitur [3][6]. Dalam prosesnya, *K-Means* menggunakan ukuran jarak *Euclidean* dan secara berulang menentukan *cluster* asal untuk setiap catatan data.

D Davies Bouldin Index

Davies Bouldin Index (DBI) adalah sebuah metrik evaluasi yang digunakan untuk menilai kualitas clustering. Metrik ini menggabungkan evaluasi terhadap sejauh mana *cluster* terbentuk dan tingkat kompaknya [8]. Dalam indeks *Davies-Bouldin*, semakin kecil nilainya menunjukkan skema klaster yang lebih baik [9].

III. METODOLOGI

Metodologi adalah langkah atau proses yang disusun secara sistematis untuk merancang dan menjalankan suatu penelitian dengan tujuan mencapai target yang diharapkan. Penelitian ini menggunakan data sekunder yaitu hanya mengolah data dari pihak lain. Penulis membuat kerangka penelitian dalam bentuk urutan gambar untuk menunjukkan tahapan dari penelitian, seperti yang terlihat pada Gambar 1.



Gambar 1. Diagram Tahapan Penelitian

A. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan mengakses dan memanfaatkan sumber data dari Badan Pusat Statistik (BPS) Sulawesi Tenggara[2]. Data yang diperoleh meliputi laporan statistik terkait jumlah kejahatan atau kriminalitas yang dilaporkan dan persentase korban kejahatan di Sulawesi Tenggara dalam periode 2021. Data tersebut dapat dilihat pada Tabel 1.

Tabel 1. Data Statistik Kriminalitas di Sulawesi Tenggara

Kabupaten/Kota	Jumlah Kejahatan	Persentase Korban
Buton	68	0.28
Muna	471	1.15
Konawe	277	1.48
Kolaka	212	0.81
Konawe Selatan	391	1.65

Bombana	53	0.58
Wakatobi	204	0.09
Kolaka Utara	67	0.68
Buton Utara	124	0.99
Konawe Utara	126	0.98
Kolaka Timur	26	3.2
Konawe Kepulauan	7	0.35
Muna Barat	81	0.87
Buton Tengah	28	0.5
Buton Selatan	30	0.05
Baubau	245	2.34
kendari	590	2.35

Langkah ini menjadi landasan untuk analisis lanjutan dalam melakukan klasterisasi menggunakan metode *K-Means Clustering* guna mengidentifikasi daerah-daerah rawan kriminalitas di Sulawesi Tenggara. Adapun penggunaan data dari BPS ini diharapkan dapat memberikan informasi yang akurat dan dapat diandalkan untuk keperluan analisis dan evaluasi dalam penelitian ini.

B. Pra-Pemrosesan Data

Praproses data berguna untuk memenuhi persyaratan dalam pengimputan *K-Means Clustering*. *K-Means Clustering* sebagai metode pengelompokan data secara khusus hanya dapat mengolah atribut yang memiliki nilai numerik. Oleh karena itu, tahapan praproses data bertujuan untuk memastikan bahwa atribut yang dipilih memiliki tipe data numerik yang dapat dianalisis oleh algoritma K-means. Adapun tahapan pra-pemrosesan data yang dilakukan sebagai berikut ini:

1) Seleksi Atribut

Dari dataset yang tersedia, dilakukan seleksi atribut untuk fokus pada variabel penting yang akan digunakan dalam analisis. Dalam penelitian ini atribut yang diperlukan adalah Jumlah Kejahatan dan Persentase Korban.

2) Normalisasi Data

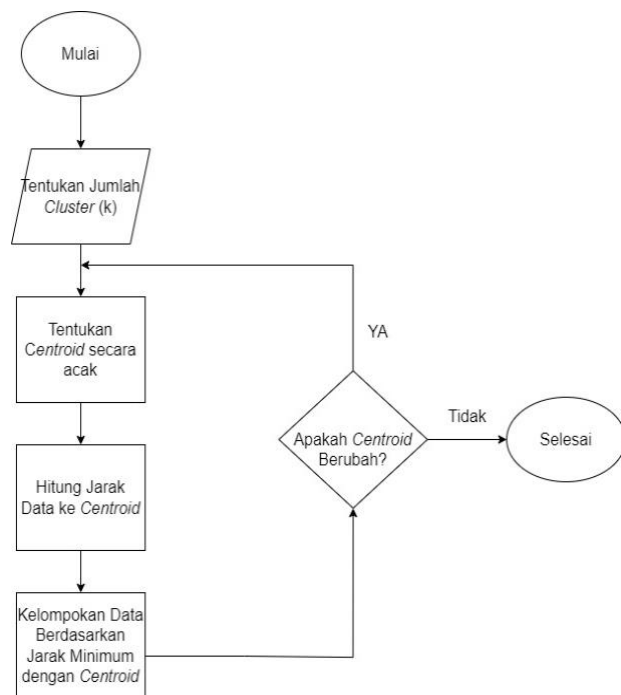
Normalisasi merupakan suatu teknik untuk mengorganisasikan data kedalam tabel-tabel untuk memenuhi kebutuhan pemakai didalam suatu organisasi. Data-data yang dilakukan normalisasi dengan membagi nilai data tersebut dengan nilai *range* data (nilai data maksimum – nilai data minimum) [5]. Pada atribut data dalam penelitian ini memiliki perbedaan skala yang signifikan antara atribut 'jumlah kejahatan' dan 'persentase korban'. Oleh karena itu, Normalisasi diperlukan untuk memastikan bahwa kedua variabel ini memiliki rentang nilai yang serupa. Ini penting agar tidak ada variabel yang mendominasi proses klasterisasi hanya karena memiliki skala nilai yang besar. Metode yang digunakan dalam normalisasi data ini ialah *MinMaxScaler*. Hasil dari normalisasi data dapat dilihat pada Tabel 2.

Tabel 2. Data Setelah dinormalisasi

Jumlah kejahatan	Persentase korban
0.10463122	0.07301587
0.7588336	0.34920635
0.46312178	0.45396825
0.3516295	0.24126984
0.65866209	0.50793651
0.07890223	0.16825397
0.33790738	0.01269841
0.10291595	0.2
0.20068611	0.2984127
0.20411664	0.2952381
0.03259005	1
0	0.0952381
0.12692967	0.26031746
0.03602058	0.14285714
0.03945111	0
0.40823328	0.72698413
1	0.73015873

C. Penerapan ritma K-Means Clustering

Pada tahapan ini akan dilakukan pengelompokan data kedalam sebuah *cluster* atau bisa disebut *clustering*[10][6]. Metode yang digunakan dalam *clustering* ini adalah *K-Means*. Adapun Langkah-langkah dalam *K-Means* adalah sebagai berikut:



Gambar 2. Diagram K-Means Clustering

- Langkah pertama adalah menentukan jumlah kluster yang optimal. Pemilihan nilai k dapat dilakukan berdasarkan kebutuhan analisis atau menggunakan metode *Elbow* untuk menemukan nilai yang optimal.
- Menentukan titik pusat kluster (*centroid*) sebanyak k akan diinisiasi secara acak. Pada tahap ini, setiap *centroid* akan merepresentasikan satu kluster.
- Menghitung jarak antara setiap data dengan masing-masing *centroid* dengan menggunakan rumus *Euclidean Distance*. Rumus *Euclidean Distance* adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Keterangan :

 d = Jarak antara x dan y

x = Data

y = *Centroid*

i = Iterasi

n = Jumlah dari data

- Kelompokkan setiap data ke dalam *cluster* berdasarkan jarak minimum dengan *centroid*. Data akan menjadi bagian dari kluster dengan *centroid* terdekat.
- Menentukan posisi *centroid* baru
- Ulangi Langkah ke 3 sampai posisi *centroid* tidak mengalami perubahan

D. Pengujian

Pada proses ini dilakukan pengujian untuk hasil dari *K-Means Clustering* sebelumnya, pengujian dilakukan dengan

cara mengevaluasi kualitas *K-Means Clustering* menggunakan metode *Davies-Bouldin Index* (DBI). Adapun rumus yang digunakan adalah seperti persamaan berikut :

$$DBI = \frac{1}{k} \sum_{i=1, i \neq j}^k \max(R_{i,j}) \quad (3)$$

Simbol k mengindikasikan jumlah kluster yang digunakan, sementara R mewakili rasio antara kluster ke- i dan ke- j , dihitung dari aspek kohesi dan pemisahan. Evaluasi kualitas kluster didasarkan pada mencari tingkat kohesi minimum dan tingkat pemisahan maksimum. Berikut ini rumus untuk mencari nilai $R_{i,j}$:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

SSB (*sum of square between cluster*) adalah metrik untuk memisahkan antara dua *cluster*. Sedangkan SSW (*Sum of Square within a cluster*) adalah metrik kohesi dalam *cluster* ke- i . Untuk mencari keduanya menggunakan rumus 5, rumus 6 dan rumus 7.

$$SSB_{i,j} = d(c_i, c_j) \quad (5)$$

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (6)$$

$$SSW_j = \frac{1}{m_j} \sum_{i=1}^{m_j} d(x_i, c_j) \quad (7)$$

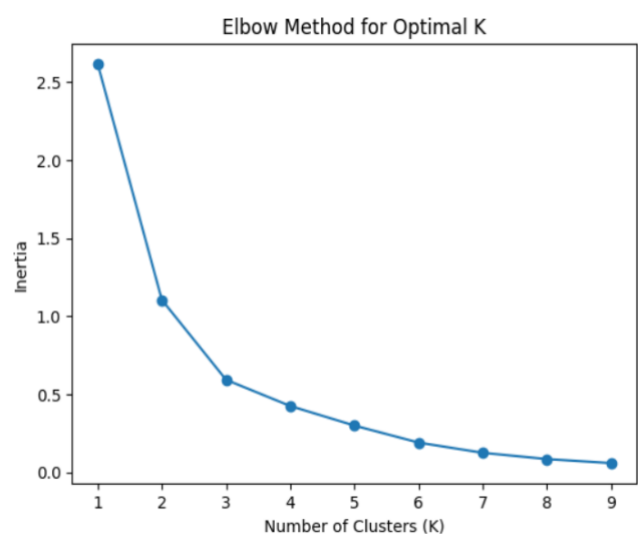
Keterangan :

 $d(x_j, c_i)$ = Jarak data ke- i ke *centroid* ke- j $d(x_i, c_j)$ = Jarak data ke- i ke *centroid* ke- j $d(c_i, c_j)$ = Jarak *centroid* ke- i ke *centroid* ke- j m_i, m_j = Jumlah data dalam *cluster* ke- i dan ke- j

IV. HASIL DAN PEMBAHASAN

A nilai k optimal

Langkah awal merupakan penentuan nilai K menggunakan metode *Elbow*, dengan tujuan memperoleh nilai K yang optimal. Grafik *Elbow* untuk nilai K optimum dapat ditemukan dalam Gambar 2.



Gambar 3. Grafik Elbow

Dapat dilihat dari grafik *Elbow* diatas bahwa nilai K yang optimum adalah 3. Hal ini dikarenakan penurunan nilai inersia pada K= 3 mulai melambat atau tidak lagi signifikan, sehingga K = 3 menjadi nilai K optimum. Dengan demikian, hasil klasterisasi akan membagi daerah menjadi tiga tingkatan rawan kriminalitas yaitu tidak Rawan (C0) , sedang (C1) dan Rawan (C2).

B Klasterisasi

Dengan nilai K dan atribut yang telah ditetapkan sebelumnya, langkah selanjutnya adalah melakukan pembagian klaster menggunakan algoritma K-Means. Implementasi proses ini dilakukan dengan menggunakan bahasa pemrograman Python. Hasil pembagian klaster ini didasarkan pada nilai K yang telah ditentukan sebelumnya, dan outputnya dapat ditemukan dalam Tabel 3.

Tabel 3. Pembagian Cluster

Kabupat en/Kota	Jumlah Kejahatan	Persenta se Korban	Cluster	Cluster label
Buton	68	0.28	0	Tidak
Muna	471	1.15	2	Rawan
Konawe	277	1.48	2	Rawan
Kolaka	212	0.81	0	Tidak
Konawe Selatan	391	1.65	2	Rawan
Bombana	53	0.58	0	Tidak
Wakatob i	204	0.09	0	Tidak
Kolaka Utara	67	0.68	0	Tidak
Buton Utara	124	0.99	0	Tidak
Konawe Utara	126	0.98	0	Tidak
Kolaka Timur	26	3.20	1	Sedang
Konawe Kepulaua n	7	0.35	0	Tidak
Muna Barat	81	0.87	0	Tidak
Buton Tengah	28	0.50	0	Tidak
Buton Selatan	30	0.05	0	Tidak
Baubau	245	2.34	1	Sedang
Kendari	590	2.35	2	Rawan

C Evaluasi Hasil Clustering

Hasil dari *clustering* menggunakan algoritma *K-Means Clustering* dapat diukur atau divalidasi jumlah *cluster* yang dihasilkan. Teknik untuk evaluasi *clustering* dilakukan dengan menggunakan *Davies Bouldin Index*. Hasil dari

evaluasi menggunakan *Davies Bouldin Index* ada pada Tabel 4.

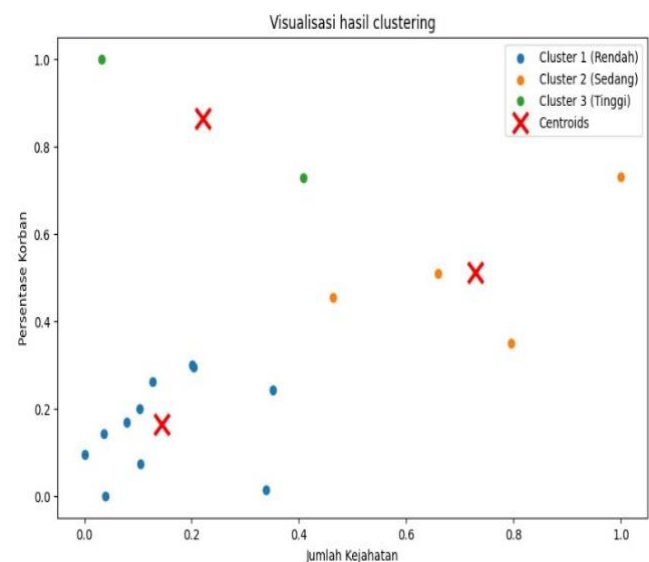
Tabel 4. Hasil Evaluasi Menggunakan Davies Bouldin Index

Cluster	SSW	SSB	R	DBI
C0	0.2536	0.5339	0.4751	0.5904
C1	0.1078	0.5859	0.1840	
C2	0.2310	0.9046	0.2553	

Dapat dilihat dari Tabel 4 untuk nilai SSW terbaik ada pada C1, nilai SSB terbaik ada pada C0, dan untuk nilai R terbaik ada pada C0. Dari Tabel 4 juga kita dapat menentukan untuk nilai *Davies Bouldin Index* pada hasil *cluster* ini adalah 0.5904.

D Visualisasi Hasil Clustering

Setelah berhasil mendapatkan hasil *clustering* optimal, langkah berikutnya melibatkan pembuatan visualisasi dari klaster yang terbentuk. Tujuan dari langkah ini adalah untuk memfasilitasi analisis terhadap masing-masing *cluster* yang dihasilkan dalam proses *clustering* sebelumnya. Gambaran visual dari hasil *clustering* dapat dilihat melalui Gambar 3.



Gambar 4. Visualisasi Hasil Clustering

V. KESIMPULAN

K-Means Clustering adalah salah satu algoritma untuk mengelompokkan suatu data, ada pun hasil dari penelitian ini didapatkan untuk *cluster* daerah rawan kriminalitas di Sulawesi Tenggara terbagi menjadi 3 yaitu daerah yang tidak rawan, sedang , dan rawan akan kriminalitas. Untuk nilai evaluasi dari hasil *clustering* ini adalah 0.5904.

DAFTAR PUSTAKA

- [1] N. Sulisrudatin, "Kasus Begal Motor Sebagai Bentuk Kriminalitas Pelajar," *Jurnal Mitra Manajemen*, hlm. 57–67, 2020, [Daring]. Tersedia pada: www.depoknews.com,

- [2] “Jumlah Kejahatan yang Dilaporkan (Crime Total) dan Persentase Tindak Kejahatan yang Diselesaikan (Clearance Rate) menurut Kepolisian Resort Tahun 2019 - 2021,” *sultra.bps.go.id*.
- [3] Orleans B dan Purnomo Putra E, “Clustering Algoritma (K-Means),” *sis.binus.ac.id*.
- [4] A. Al Fahrozi, F. Insani, E. Budianita, dan I. Afrianty, “Implementasi Algoritma K-Means dalam Menentukan Clustering pada Penilaian Kepuasan Pelanggan di Badan Pelatihan Kesehatan Pekanbaru,” *Indonesian Journal of Innovation Multidisipliner Research*, vol. 1, no. 4, hlm. 474–492, 2023.
- [5] C. C. I. Loho, V. P. Rantung, dan G. C. Rorimpandey, “Data Mining Rekomendasi Sekolah Calon Siswa SMA di Kota Tomohon Menggunakan Metode K-Means Clustering,” *JOINTER : Journal Of Informatics Engineering*, vol. 03, no. 01, hlm. 30–38, 2022.
- [6] A. Aditya, I. Jovian, dan B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019,” *Jurnal Media Informatika Budidarma*, vol. 4, no. 1, hlm. 51, Jan 2020, doi: 10.30865/mib.v4i1.1784.
- [7] S. Adiningsi dkk., “Klasifikasi Bentuk Telur Ayam Berbasis Citra Digital Menggunakan Metode K-Means Clustering,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 3, hlm. 1507–1512, 2023.
- [8] R. Ishak dan Amiruddin, “Penerapan Algoritma XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah,” *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, hlm. 192–196, 2023.
- [9] N. Azmi, H. HS, Yuyun, dan Hazriani, “Penerapan Metode K-Means Clustering Dalam Mengelompokkan Data Penjualan Obat pada Apotek M23,” *SISFOTEK*, hlm. 244–248, 2023.
- [10] S. Setiawan, “Pemanfaatan Metode K-Means Dalam Penentuan Persediaan Barang,” *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, vol. 6, no. 1, hlm. 41–48, 2018.